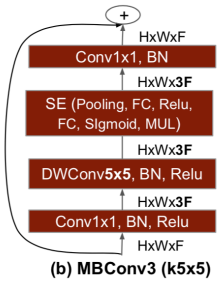
ML2020SPRING HW3 Report

學號：R08946015 系級：資料科學碩一 姓名：陳鈞廷

1. 請說明你實作的 CNN 模型，其模型架構、訓練參數量和準確率為何?

這次我實作的 CNN 模型參考了 EfficientNet (Tan, M., & Le, Q. V. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:1905.11946.) 的架構，EfficientNet 的架構主要是連接數個 MBConv 的 block 而組成，每個 MBConv 首先會有一層 pointwise convolution 來擴張 channel 數量，接著連接一層 deepwise convolution 來抽取特徵，然後再連接 Squeeze-and-excitation (Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7132-7141)，最後連結一層 pointwise convolution 將 channel 數量縮減回 input 的 channel 數量。下圖為 MBConv 的簡單示意圖。



(b) MBConv3 (k5x5)

經過了數次實驗後，我發現增加 MBConv 的數量（深度）並沒有明顯的 performance 提升，因此我只擴張 model 的 width（channe 數量），最後經過調整後，input image size 為 300 x 300，模型總共有 15 個 MBConv 的 block，總參數量為 5,808,769，所有參數皆為 trainable。

在模型的訓練 optimizer，我使用原本論文中推薦的 RMSProp，初始 learning rate 設定為 0.0005，learning rate 每過 3 個 epochs 會縮減成 0.97 倍，總共訓練 120 個 epochs，挑選 validation acc. 最高的，validation acc. 達到 0.8460，而 kaggle 的 public score 達到 0.85056。

2. 請實作與第一題接近的參數量，但 CNN 深度(CNN 層數)減半的模型，並說明其模型架構、訓練參數量和準確率為何?

這裡我參考助教提供的 example code 中的 CNN 架構，總共 5 層 convolution，但縮減了最後兩層 convolution 的 channel 數量，5 層 convolution 的 channel 數依序為 64、128、256、256、256，最後參數量達到 6,278,667，所有參數皆為 trainable。。

在模型的 optimizer 上我選擇了 Adam，初始 learning rate 設定為 0.001，並套用一樣的 learning rate decay，訓練 100 個 epoch 後 validation accuracy 最高達到 0.75。

3. 請實作與第一題接近的參數量，簡單的 DNN 模型，同時也說明其模型架構、訓練參數和準確率為何?

我實作了 3 層 linear 的簡單 DNN 模型，3 層的 unit 數量分別為 128、256、128，每層 linear 後面連接 batch normalization 1d 和 ReLU，總參數量達到 6,359,947，所有參數皆為 trainable。模型訓練的方法跟上題 CNN 的一樣，在訓練 100 個 epoch 後 validation accuracy 最高達到 0.38。

4. 請說明由 1 ~ 3 題的實驗中你觀察到了什麼?

從分數上的差異異觀察，可以發現簡單 CNN 的表現比 DNN 好上非常多，我認為雖然 CNN 和 DNN 的參參數量量差不多，但由於 CNN 的 convolution layer 可以讓參參數量量減少，使得 CNN 可以在同參參數量量的限制下擁有較多層 layer，因此可以提取出較複雜的 feature，所以 CNN 的表現才會比 DNN 好。

而我實作的類 EfficientNet 的準確率比簡單 CNN 高出大約 0.09，因此可以發現深度加深對於準確率是可以提升的，然而 EfficientNet 在訓練的過程中 validation loss 並非穩健的下降，就連 validation accuracy 也會出現明顯的 oscillation，因此我認為較深的 CNN 雖然 performance 較好，但訓練較不容易。

5. 請嘗試 data normalization 及 data augmentation，說明實作方法並且說明實行前後對準確率有什麼樣的影響?

首先我計算出所有 train set 圖片的 RGB 三個 channel 的 mean 和 standard deviation，data normalization 是扣掉 mean 在除以 standard deviation，其中 mean = [0.55474155, 0.45078358, 0.34352523]，std [0.2719837, 0.27492649, 0.28205909]。

Data augmentation 使用 torchvision 提供的 transforms，我使用了 RandomHorizontalFlip、RandomResizedCrop、RandomAffine，其中 RandomResizedCrop 設定了縮放比例從 0.8 到 1.2，aspect 比例從 0.8 到 1.2；RandomAffine 中設定 30º 的旋轉、0.2 的平移、縮放範圍 0.9 ~ 1.15、以及 15º 的 shear。

下表為實行前後的 validation 正確率比較，可以發現 data normalization 對準確率影響較小，可能是因為原本已經將 pixel 值除以 255 了，所以變動不大。另外可以發現 Data augmentation 對於準確率有較高的影響。

	Validation Accuarcy
No	0.78
Data Augmentation	0.82
Data Normalization	0.79
Data Aug. + Normalization	0.84

6. 觀察答錯的圖片中，哪些 class 彼此間容易用混?[繪出 confusion matrix 分析]

下圖為 confusion matrix，可以發現 Dairy product 的正確率最低，而且 Dairy product 有 0.174 的比例被誤判成 Dessert。